

IMMEDIATE COMMUNICATION

A rare missense mutation in *CHRNA4* associates with smoking behavior and its consequences

TE Thorgeirsson¹, S Steinberg¹, GW Reginsson¹, G Bjornsdottir¹, T Rafnar¹, I Jonsdottir^{1,2}, A Helgadottir¹, S Gretarsdottir¹, H Helgadottir¹, S Jonsson^{2,3}, SE Matthiasson⁴, T Gislason^{2,3}, T Tyrfinngsson⁵, T Gudbjartsson^{2,6}, HJ Isaksson⁷, H Hardardottir³, A Sigvaldason³, LA Kiemenev^{8,9}, A Haugen¹⁰, S Zienolddiny¹⁰, HJ Wolf¹¹, WA Franklin¹², A Panadero¹³, JI Mayordomo¹⁴, IP Hall¹⁵, E Rönmark^{16,17}, B Lundbäck^{16,18}, A Dirksen¹⁹, H Ashraf^{19,20}, JH Pedersen²¹, G Masson¹, P Sulem¹, U Thorsteinsdottir¹, DF Gudbjartsson¹ and K Stefansson^{1,2}

Using Icelandic whole-genome sequence data and an imputation approach we searched for rare sequence variants in *CHRNA4* and tested them for association with nicotine dependence. We show that carriers of a rare missense variant (allele frequency = 0.24%) within *CHRNA4*, encoding an R336C substitution, have greater risk of nicotine addiction than non-carriers as assessed by the Fagerstrom Test for Nicotine Dependence ($P = 1.2 \times 10^{-4}$). The variant also confers risk of several serious smoking-related diseases previously shown to be associated with the D398N substitution in *CHRNA5*. We observed odds ratios (ORs) of 1.7–2.3 for lung cancer (LC; $P = 4.0 \times 10^{-4}$), chronic obstructive pulmonary disease (COPD; $P = 9.3 \times 10^{-4}$), peripheral artery disease (PAD; $P = 0.090$) and abdominal aortic aneurysms (AAAs; $P = 0.12$), and the variant associates strongly with the early-onset forms of LC (OR = 4.49, $P = 2.2 \times 10^{-4}$), COPD (OR = 3.22, $P = 2.9 \times 10^{-4}$), PAD (OR = 3.47, $P = 9.2 \times 10^{-3}$) and AAA (OR = 6.44, $P = 6.3 \times 10^{-3}$). Joint analysis of the four smoking-related diseases reveals significant association ($P = 6.8 \times 10^{-5}$), particularly for early-onset cases ($P = 2.1 \times 10^{-7}$). Our results are in agreement with functional studies showing that the human $\alpha 4\beta 2$ isoform of the channel containing R336C has less sensitivity for its agonists than the wild-type form following nicotine incubation.

Molecular Psychiatry (2016) **21**, 594–600; doi:10.1038/mp.2016.13; published online 8 March 2016

INTRODUCTION

Smoking is a risk factor for many serious diseases and a leading cause of preventable deaths worldwide.¹ Genome-wide association studies relying on common markers have yielded loci associating with smoking behavior, including genomic regions containing genes encoding various subunits of nicotinic acetylcholine receptors (nAChRs) on chromosomes 15q25 (*CHRNA5/CHRNA3/CHRNA4*),^{2,3} 8p11 (*CHRNA3/CHRNA6*)³ and recently 20q11 (*CHRNA4*).⁴ Sequence variants within the *CHRNA5/CHRNA3/CHRNA4* cluster on chromosome 15q25 have been shown to associate with the number of cigarettes smoked per day (CPD),^{2,5} nicotine dependence (ND)^{2,6} and the smoking-related diseases lung cancer (LC),^{2,7–9} peripheral arterial disease (PAD),² chronic obstructive pulmonary disease (COPD)¹⁰ and upper aerodigestive tract cancer.¹¹ The role of the *CHRNA5/CHRNA3/CHRNA4* locus for smoking, ND and the consequences of smoking was established through human genetics approaches,

and these initial findings have paved the way for basic research aimed at elucidating the underlying mechanisms.¹²

nAChR channels are pentameric and comprised of various types of subunits. Twelve α subunits and three β subunits have been cloned, and various isoforms containing both $\alpha 4$ and $\beta 2$ subunits are the most abundant in mammalian brain.¹³ nAChRs have a large dynamic range with respect to agonist activation, but this is partially explained by the presence of heteromers with different stoichiometric composition of subunits with different sensitivities. Measurements indicate that $(\alpha 4)_3(\beta 2)_2$ is a low-sensitivity subtype (LS), whereas $(\alpha 4)_2(\beta 2)_3$ is of high sensitivity (HS).¹⁴ Chronic nicotine exposure upregulates nAChRs in a complex process involving molecular chaperoning,^{15,16} and upregulation of $\alpha 4\beta 2$ receptors by nicotine favors the HS form.¹⁷ Lowered sensitivity to psychoactive drugs is generally believed to correlate with increased risk of substance dependence, and it has been suggested that increased expression of LS relative to HS nAChRs in the central nervous system would increase risk of ND.¹⁸

¹deCODE genetics/Amgen, Reykjavik, Iceland; ²Faculty of Medicine, University of Iceland, Reykjavik, Iceland; ³Department of Respiratory Medicine, Landspítali University Hospital, Reykjavik, Iceland; ⁴Laekning, Medical Clinics, Reykjavik, Iceland; ⁵SAA National Center of Addiction Medicine, Reykjavik, Iceland; ⁶Department of Cardiothoracic Surgery, Landspítali University Hospital, Reykjavik, Iceland; ⁷Department of Pathology, Landspítali University Hospital, Reykjavik, Iceland; ⁸Radboud Institute for Health Sciences, Radboud University Medical Center, Nijmegen, The Netherlands; ⁹Department of Urology, Radboud University Medical Center, Nijmegen, The Netherlands; ¹⁰Department for the Chemical and Biological Work Environment, National Institute of Occupational Health, Oslo, Norway; ¹¹Community & Behavioral Health, Colorado School of Public Health, University of Colorado Denver, Aurora, CO, USA; ¹²Department of Pathology, University of Colorado Denver, Aurora, CO, USA; ¹³Division of Medical Oncology, Hospital Ciudad de Coria, Coria, Spain; ¹⁴Division of Medical Oncology, University of Colorado School of Medicine, Denver, CO, USA; ¹⁵Division of Respiratory Medicine, Queens Medical Centre, University of Nottingham, Nottingham, UK; ¹⁶The OLIN studies, Department of Medicine, Sunderby Central Hospital of Norrbotten, Luleå, Sweden; ¹⁷Faculty of Medicine, Department of Public Health and Clinical Medicine, Occupational and Environmental Medicine, Umeå University, Umeå, Sweden; ¹⁸Krefting Research Centre, Institute of Medicine, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; ¹⁹Department of Respiratory Medicine, Gentofte Hospital, Copenhagen University, Hellerup, Denmark; ²⁰Centre for Diagnostic Imaging—Thoracic Section, Akershus University Hospital, Loerenskog, Norway and ²¹Department of Thoracic Surgery RT, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark. Correspondence: Dr TE Thorgeirsson or K Stefansson, deCODE genetics/Amgen, Sturlugata 8, Reykjavik IS-101, Iceland.

E-mail: thorgeir@decode.is or kstefans@decode.is

Received 22 January 2015; revised 17 December 2015; accepted 4 January 2016; published online 8 March 2016

Table 1. Association of eight rare missense variants in *CHRNA4* with FTND

Variant	P-value	Effect	MAF (%)	Imp Inf	Position	Minor allele	Major allele	Coding effect
chr20:61451525:S	0.37	-0.71	0.017	0.88	61 451 525	A	G	S561L
rs77345643	0.099	0.91	0.038	0.73	61 451 540	A	G	P556L
rs55915440	0.48	0.063	1.378	0.96	61 451 855	A	G	P451L
rs121912272	0.76	-0.042	0.547	0.91	61 451 979	T	C	V410I
rs56175056	1.2×10^{-4}	0.85	0.237	0.99	61 452 201	A	G	R336C
chr20:61457796:S	0.15	0.41	0.127	0.96	61 457 796	A	G	R120W
chr20:61457871:S	0.68	0.28	0.05	0.44	61 457 871	G	C	D95H
chr20:61461433:S	0.013	0.61	0.61	0.97	61 461 433	T	C	G47S

Abbreviations: FTND, Fagerström Test of Nicotine Dependence; Imp, imputation; Inf, information; MAF, minor allele frequency. Table 1 shows association results for non-synonymous variants in *CHRNA4* identified by sequencing 2636 Icelanders and imputing their variants into 104,220 Icelanders for the FTND run as a quantitative trait using three categories (FTND 0–3 = mild, FTND 4–6 = moderate and FTND 7–10 = severe). Shown are identity of variant (rs names where available), P-value of association, effect size, minor allele frequency in Iceland (in %), imputation information, position of each variant on chromosome 20 in build 36, identity of minor and major alleles, and the coding effect of the variant. The reported effect size is in FTND units, obtained by multiplying the effect size from analysis with 3.5, the difference between midpoints of adjacent categories.

Results of animal studies indicate that both $\alpha 4$ and $\beta 2$ subunits influence key addiction-related processes.¹³ *CHRNA4* and *CHRNA2* have long been considered among the strongest functional candidate genes in the genome when it comes to nicotine addiction and related phenotypes, and an efficacious smoking cessation aid (varenicline) was developed as a partial agonist of the $\alpha 4\beta 2$ isoforms.¹⁹ A large number of candidate gene studies have targeted *CHRNA4* and *CHRNA2*, but the results have been inconclusive. In a companion paper⁴ we show that common markers within *CHRNA4* exhibit genome-wide significant association with score on the Fagerström Test of Nicotine Dependence (FTND),²⁰ the most widely used measure of nicotine addiction. The genome-wide association study of FTND was limited to the study of common variants,⁴ and here we study rare variants within *CHRNA4*, by testing eight missense variants in *CHRNA4* for association with FTND.

MATERIALS AND METHODS

Study populations

The Icelandic studies include a number of ongoing projects in Iceland. All Icelandic studies were approved by the Icelandic Data Protection Authority and the National Bioethics Committee. All subjects who donated samples also signed informed consent. Personal identifiers of the patients and biological samples were encrypted by a third-party system provided by the Icelandic Data Protection Authority. In addition, replication studies were performed by genotyping samples from a number of studies conducted elsewhere. All subjects in these studies gave written informed consents and are of European origin. The association studies of the variants were performed using genotypes from whole-genome sequence data and imputation approaches previously outlined.²¹ Further details of the various studies involved are provided below (also see Results and Tables).

Smoking phenotypes

The discovery sample utilized FTND data from questionnaires answered by participants in deCODE's study of ND.^{2,3} Responses to FTND questions generate a score of 0–10, with higher scores representing greater ND.²⁰ Information on smoking quantity (SQ) was also utilized, but SQ was available from a standardized smoking questionnaire used in deCODE's studies that asks: 'How many cigarettes per day do/did you smoke on average (on most days)?' This means that current smokers answer their current consumption and former smokers refer to their consumption in the past. In cases where multiple records were available we recorded the maximum. The SQ was categorized into four levels,

(1–10, 11–20, 21–30 and 31+ CPD). A total of 40 573 subjects (34 200 chip-typed) were included in the R336C analysis. Ever smokers were recruited in the years 1997–2014 as part of various Icelandic studies, and the FTND questionnaires were administered to participants in the study of ND (2004–2014). Analysis of subjects with complete FTND data showed that of those reporting 1–10 CPD, 85% had a total score of 3 or less on the FTND, and < 3% scored above 6 on the FTND. We included 4313 chip-typed low-quantity smokers (CPD < 11) but without additional FTND items answered to the FTND 0–3 group for the meta analysis of common marker association with FTND.⁴ In our studies of R336C we included additional low-quantity smokers in this category, bringing the total number in the category to 18 184, with 15 527 chip-typed. The number of subjects per category were for R336C analysis:

FTND 0–3 (mild ND): 18 184 (15 527 chip-typed)
 FTND 4–6 (moderate ND): 2008 (1980 chip-typed)
 FTND 7–10 (severe ND): 1106 (1085 chip-typed).

Smoking-related diseases

Lung cancer

Iceland: The Icelandic LC study has been described previously.² The primary source of information on the Icelandic LC cases is the Icelandic Cancer Registry, which covers the entire population of Iceland (<http://www.cancerregistry.is>). Briefly, according to the Icelandic Cancer Registry, a total of 4560 LC patients were diagnosed from 1 January 1955 to 31 December 2012. Recruitment of both prevalent and incident cases was initiated in 1998, the recruitment is ongoing and DNA samples from LC cases were subjected to whole-genome genotyping as they are collected. The LC associations were based on 1437 chip-typed LC cases (median age 70 years, range 16–97, 48% males) and 84 086 chip-typed controls in addition to 2572 LC cases and 137 443 controls who had at least one of their first- or second-degree relatives chip-typed. To study early-onset LC we considered subjects diagnosed before reaching an age of 58 years.

Nijmegen, The Netherlands: The Dutch study population was previously described.²² The 634 patients with LC were identified through the population-based cancer registry of the Comprehensive Cancer Center IKO, Nijmegen, the Netherlands, and recruited through several independent studies.²² The 4670 cancer-free controls were selected from participants of the 'Nijmegen Biomedical Study'.²³ All controls are of self-reported European descent. The study protocols of the Nijmegen Biomedical Study were approved by the Institutional Review Board of the Radboud University Nijmegen Medical Centre.

Zaragoza, Spain: The Spanish study population was previously described.²² The 455 LC cases were recruited from the Oncology Department of Zaragoza Hospital in Zaragoza, from June 2006 to June 2009. The 1444 Spanish controls were patients at the University Hospital in Zaragoza for diseases other than cancer, between November 2001 and May 2007. Study protocols were approved by the Institutional Review Board of Zaragoza University Hospital.

Denver, Colorado: DNA samples from blood and clinical data were provided by the University of Colorado Cancer Center under COMIRB protocol 08–0380. Blood samples were collected from 1217 patients diagnosed with different diseases enrolled in any of 20 clinical research trials carried out at Colorado SPORE protocols between 1993 and 2008. Of these 1217 patients, 246 were LC cases and 971 had never had LC at the time of sample shipment. LC cases were identified either from data matches with the Colorado Central Cancer Registry or by having malignant lung tissue collected via enrollment in a surgical protocol. Work in this study was limited to cases and controls of self-reported European ancestry (195 cases and 798 controls).

Norway: Comprehensive details of cases and controls have been previously published.²⁴ Briefly, a systematic series of 426 patients with histologically proven non-small cell LC treated with surgery were ascertained through university hospitals in Oslo or Bergen between 1986 and 2001. All of the cases were either smokers at the time of surgery or ex-smokers. Healthy controls were recruited from a general health survey of individuals conducted by the National Health Surveys in the Oslo area (HUBRO) between 2000 and 2001, or in western Norway. Among them, smokers without any known history of cancer were randomly selected and frequency matched to the cases on age, smoking dose (pack-years) and gender. Cases and controls were interviewed using questionnaires that allowed collection of the same demographic and lifestyle information. Both cases and controls were Norwegians of European descent. The study was approved by the regional ethical committee in accordance with the Helsinki Declaration.

Chronic obstructive pulmonary disease

Iceland: COPD was defined broadly and included diagnoses of chronic bronchitis or emphysema made between 1980 and 2011 at the Landspítali University Hospital. A total of 8246 subjects born between 1883 and 1964 were included, with most (5871) of the confirmed COPD diagnoses made during the years 1998–2011. The actual age at onset of COPD is difficult to assess, as patients have usually had signs and symptoms for a long time before admission to hospital. To define an early-onset group we used year of birth as a surrogate, using the group born between 1940 and 1964 ($N = 1470$) representing the youngest 19.6% of patients. Association analyses were based on 3857 chip-typed cases, and 3639 patients with at least one first- or second-degree relative chip-typed.

United Kingdom: A total of 206 subjects with COPD defined by spirometry (forced expiratory volume (FEV)/forced vital capacity (FVC) < 0.7 , FEV1% $< 80\%$), over 40 years of age, and 338 controls, as part of the Nottingham Smokers Study.²⁵

Sweden: Participants in a population study on COPD in Norrbotten (Olin study), Sweden.²⁶ Those who answered questionnaire were followed up with spirometry and blood sample; 581 COPD patients (defined by spirometry (FEV1/FVC < 0.7 and FEV1 $< 80\%$ of predicted values) or (FEV1/FVC < 0.7)) and 918 controls.

Denmark: Participants in a LC screening study of 4104 current or previous smokers (2004–2006),²⁷ 708 were diagnosed with COPD based on spirometry (FEV1/FVC < 0.7) and emphysema based on high resolution computed tomography scan.

Peripheral arterial disease

Iceland: The study population has been described previously.² Briefly, Icelandic individuals with PAD were recruited from a

registry of individuals diagnosed with PAD at the Landspítali University Hospital, during the years 1983–2011. The PAD diagnosis was confirmed by vascular imaging or segmental pressure measurements. To study early-onset PAD we considered subjects diagnosed with PAD before reaching age 65.

Abdominal aortic aneurysm

Iceland: Icelanders with abdominal aortic aneurysms (AAAs) were recruited from a registry of individuals who were admitted either for emergency repair of symptomatic or ruptured AAA, or for an elective surgery to the Landspítali University Hospital, in Reykjavik, Iceland, in 1980–2011. Individuals (with an unruptured aneurysm ≥ 30 mm) were also recruited through a high-risk screening of first-degree relatives of AAA patients. To study early-onset AAA we considered subjects diagnosed with AAA before reaching age 65.

Single-track assay single-nucleotide polymorphism genotyping: Single single-nucleotide polymorphism (SNP) genotyping was carried out by deCODE Genetics in Reykjavik, Iceland, applying the Centaurus (Nanogen) platform.⁷ The same assay was used to obtain genotypes for rs56175056 on three COPD sample sets, from the United Kingdom, Sweden and Denmark, as well as LC cases and controls from the Netherlands, Spain, the United States and Norway.

Illumina SNP chip genotyping: The Icelandic chip-typed samples were assayed with the Illumina HumanHap300, HumanCNV370, HumanHap610, HumanHap1M, HumanHap660, Omni-1, Omni 2.5 or Omni Express bead chips (Illumina, San Diego, CA, USA) at deCODE genetics. SNPs were excluded if they had (i) yield lower than 95%, (ii) minor allele frequency $< 1\%$ in the population, (iii) significant deviation from Hardy–Weinberg equilibrium in the controls ($P < 0.001$), (iv) if they produced an excessive inheritance error rate (over 0.001) or (v) if there was substantial difference in allele frequency between chip types (from just a single chip if the problem that resolved all differences, but from all chips otherwise). All samples with a call rate below 97% were excluded from the analysis. For the HumanHap series of chips, 308 840 SNPs were used for long-range phasing, whereas for the Omni series of chips 642 079 SNPs were included.

Whole-genome sequencing and SNP calling: SNPs were imputed based on whole-genome sequence data from about 2636 Icelanders, selected for various neoplastic, cardiovascular, metabolic and psychiatric conditions. The sample preparation, DNA-sequencing methodology, alignment and BAM file generation have been described previously.²¹

Genotype imputation methods

Long-range phasing and genotype imputation: Long-range phasing of all chip-genotyped individuals was performed with methods described previously,²¹ and imputed into chip-typed individuals and their close relatives using methods²¹ based on IMPUTE.²⁸ In brief, phasing is achieved using an iterative algorithm, which phases a single proband at a time given the available phasing information about everyone else that shares a long haplotype identically by state with the proband. Given the large fraction of the Icelandic population that has been chip-typed, accurate long-range phasing is available genome wide for all chip-typed Icelanders. The informativeness of genotype imputation was estimated by the ratio of the variance of imputed expected allele counts and the variance of the actual allele counts:

$$I = \text{Var}[E(\theta|\text{chipdata})]/\text{Var}(\theta)$$

where $\theta \in \{0,1\}$ is the allele count. $\text{Var}[E(\theta|\text{chipdata})]$ was estimated by the observed variance of the imputed expected counts and $\text{Var}(\theta)$ was estimated by $p(1-p)$, where p is the allele frequency. For

Table 2. *CHRNA4* R336C^a confers risk of heavy smoking and smoking-related diseases

Phenotype	Eff/OR	P-value	Cases		Controls	
			N chip	N rel	N chip	N rel
FTND score (mild, moderate, severe)	0.85 ^b	1.2×10^{-4}	18 592	2706	—	—
CPD in smokers	3.0 CPD	1.8×10^{-4}	34 200	6373	—	—
HSI score 4+ vs low-quantity smokers ^c	4.77	2.1×10^{-5}	1713	31	13 948	1,064
LC	2.29	4.0×10^{-4}	1437	2572	84 086	137 443
Early-onset LC	4.49	2.2×10^{-4}	245	583	85 273	129 010
PAD	1.66	0.090	1926	961	81 992	109 548
Early-onset PAD	3.47	9.2×10^{-3}	447	81	70 988	69 570
COPD	1.91	9.3×10^{-4}	3857	3639	87 510	207 914
Early-onset COPD	3.22	2.9×10^{-4}	1033	437	64 979	73 138
AAAs	1.98	0.12	574	686	83 349	122 962
Early-onset AAA	6.44	6.3×10^{-3}	92	112	83 831	123 536
All smoking-related diseases	1.87	6.8×10^{-5}	6594	6910	84 776	225 225
All smoking-related diseases early onset	3.48	2.1×10^{-7}	1704	1178	83 818	138 837

Abbreviations: AAA, abdominal aortic aneurysm; COPD, chronic obstructive pulmonary disease; CPD, cigarettes smoked per day; FTND, Fagerström Test of Nicotine Dependence; HSI, heavy smoking index; LC, lung cancer; OR, odds ratio; PAD, peripheral artery disease; SNP, single-nucleotide polymorphism. Table 2 shows the results of association analyses of two quantitative traits, nicotine dependence (FTND score) and smoking quantity (CPD), and case-control comparisons for a number of smoking-related diseases. Shown are, for each of the phenotypes listed, the effect sizes (Eff) for quantitative traits or ORs for case-control comparisons, the *P*-value and for both cases and controls, the number of subjects included in analysis, chip-typed individuals (*N* chip) and those with at least one first- or second-degree relative typed (*N* rel). The inflation factors for the various analyses listed are provided in the Supplementary Material. ^aThe SNP encoding R336C (rs56175076) has position 61 452 201 on chromosome 20 according to build 36 (hg18). ^bEffect size from analysis multiplied by 3.5, the difference in FTND score between the category midpoints; for information on the FTND analysis see Materials and Methods. ^cHSI²⁹ vs smokers smoking 10 CPD or less.

further information on the imputation methodology see previous descriptions of our methods.²¹

Association analysis. The details of our methodology for association analysis have been described,²¹ but a brief description is provided below.

Case:control association testing: Logistic regression was used to test for association between SNPs and disease, treating disease status as the response and expected genotype counts from imputation or allele counts from direct genotyping as predictor. Testing was performed using the likelihood ratio statistic. When testing for association using the imputed genotypes, controls were matched to cases based on the informativeness of the imputed genotypes, such that for each case controls of matching informativeness were chosen. Failing to match cases and controls will lead to a highly inflated genomic control factor, and in some cases may lead to spurious false-positive findings.

Quantitative traits: A generalized form of linear regression was used to test the correlation between the variants tested and quantitative traits (FTND and CPD) in Iceland. The generalized form assumes that the smoking behavior of related individuals is correlated proportional to the kinship between them rather than assuming that the smoking phenotypes of all individuals are independent.

Inflation factor adjustment: In order to account for the relatedness and stratification within the case and control sample sets we applied the method of genomic control based on chip-typed markers using a subset of about 300 000 common variants. Quoted *P*-values were adjusted accordingly. The inflation factors for the various analyses are listed in the Supplementary Material.

Analysis of replication data: Association analysis for each replication study group was carried out using Fisher's exact test, and results for different study groups were combined using the Mantel-Haenszel exact test.

RESULTS

To search for rare variants likely to influence function in *CHRNA4* we analyzed whole-genome sequence data from 2636 Icelandic subjects, identifying eight non-synonymous variants present in

high enough frequency to allow imputation into our long-range phased data set²¹ (usually approximately five carriers in our sequence data, currently corresponding to an allele frequency of about 0.1%), generating genotypes for 104 220 chip-typed Icelanders and 294 212 close relatives. Eight variants met our criteria, and following imputation they were tested for association with the FTND score (Table 1) with the significance threshold set at $P=6.25 \times 10^{-3}$ (0.05/8). The mutation encoding R336C associates with FTND score ($P=1.2 \times 10^{-4}$), whereas none of the other mutations tested showed significant association (Table 1). To improve imputation accuracy, we typed 1441 subjects for the R336C variant using a single-marker assay (Nanogen) and repeated the imputation process. As a result of this effort the imputation information increased from 0.939 to 0.999, and all results reported here were based on re-imputed data. On the basis of 1408 (original imputation) or 1441 individuals (re-imputation) with both typed and imputed genotypes the imputation accuracy for this variant as estimated by the correlation with typed genotypes changed from 0.946 (original imputation) to 0.997 (re-imputation).

There is some evidence for a founder effect as the frequency of R336C (0.24%) is higher in the Icelandic population than in any other population with a reported frequency. In the ExAC Browser (www.exac.broadinstitute.org), 31 carriers of R336C are reported in non-Finnish European populations, indicating an allele frequency of 0.046%. The variant was not detected in Asian or Latino populations, and 2 carriers from African populations ($f=0.019\%$) are reported.

To explore this association further we also examined association of the variant coding for R336C substitution with several sub-phenotypes related to nicotine addiction and the health consequences of smoking (Table 2). In selecting the smoking phenotypes and smoking-related diseases to study we were guided by previously observed associations between the key variant in *CHRNA5*, rs1051730/rs16969968 and CPD, LC, PAD, COPD and upper aerodigestive tract cancers.¹² In addition, we have observed association of rs1051730-A with AAAs in Iceland (odds ratio (OR)=1.19, $P=0.009$). As a subtype of ND we also include the heavy smoking index,²⁹ a measure that

Table 3. Results of replication studies for R336C

Sample	P-value	OR	N cases	AF cases	N ctrls	AF ctrls	Carriers	
							Cases	Ctrls
<i>LC</i>								
The Netherlands	0.32	3.7	634	7.9×10^{-4}	4670	2.1×10^{-4}	1	2
Spain	1	0	455	0	1444	3.5×10^{-4}	0	1
USA	0.10	8.2	195	5.1×10^{-3}	798	6.3×10^{-4}	2	1
Norway	—	—	426	0	758	0	0	0
All LC	0.098	3.8	1710	—	7670	—		
<i>COPD</i>								
UK	1	1.6	206	2.4×10^{-3}	338	1.5×10^{-3}	1	1
Sweden	—	—	581	0	918	0	0	0
Denmark	1	0	708	0	924	5.4×10^{-4}	0	1
All COPD	1	0.8	1495	—	2180	—		

Abbreviations: COPD, chronic obstructive pulmonary disease; LC, lung cancer; OR, odds ratio. Table 3 shows the results from replication studies of the R336C in populations outside of Iceland. Listed are for each sample set: P-value, OR, number of cases and controls (N cases and N ctrls), allele frequencies (AF cases and AF ctrls) and the number of observed carriers for cases and controls.

combines two of the key questions on the FTND,²⁰ SQ and the time to first cigarette after waking up, with a score of 4 or higher being considered high and comparable to an FTND score of ~6.³⁰ For the smoking-related diseases we also considered the effect of the variant in early-onset cases separately (Materials and Methods).

All the smoking-related diseases with observed associations with rs1051730/rs16969968 show some evidence of association with R336C (Table 2), and the effects are larger for early-onset groups. Many of the phenotypes tested are correlated, such that the individual associations with the different smoking-related diseases cannot be considered independent. To assess the overall significance of the effect that the variant has on the smoking-related diseases we combined the lists and assessed the whole group for association with R336C. This is justified by the fact that the difference between the smoking-related diseases can be looked upon as variation in expressivity of one environmental factor, namely tobacco smoke. The overall P-values obtained from this analysis are 6.8×10^{-5} and 2.1×10^{-7} for the total and early-onset smoking-related diseases, respectively, providing an estimate of the combined strength of the evidence for association with the health consequences of smoking (Table 2).

Although the associations with smoking-related diseases do not represent a fully independent replication of the FTND result, we note that because we often lack detailed information on smoking for subjects with smoking-related diseases there is very little overlap between the group of subjects with FTND data and those with smoking-related diseases. For example, only 140 subjects are both on the early-onset smoking-related disorders list and among the subjects scoring 4 or higher on the heavy smoking index, representing ~8% of the heavy smoking index cases and 4.7% of the subjects with early-onset smoking-related diseases. The overlap between these lists is so small because only a few of the patients with early-onset smoking-related disorders have filled out the FTND questionnaire. Hence, despite the correlation between smoking behavior and the smoking-related diseases our results for FTND and the smoking-related diseases are not due to direct confounding as there is very little overlap between the study groups. Furthermore, as was the case for *CHRNA5* (refs 2,31,32) the observed effect on FTND and CPD alone is not expected to fully explain the observed OR for smoking-related diseases.

The associations between the R336C variant and the various smoking-related phenotypes are not explained by

associations with common markers nearby, nor does the variant explain the association with rs2273500, the SNP most strongly associated in the meta analysis study of FTND,⁴ as the rare missense variant is fixated on the protective background of rs2273500. For the results of detailed conditional analyses see Supplementary Material.

We attempted replication studies of the findings for the R336C variant, focusing on LC (1710 cases and 7670 controls) and COPD (1495 cases and 2180 controls), using a single SNP assay for genotyping (Table 3). Combining both phenotypes, the OR was 2.3 ($P=0.25$). The low allele frequencies in these populations mean that we lack power to obtain meaningful results with the sample sizes studied, with only 10 carriers observed in 13 055 samples (average allele frequency of 0.038%).

DISCUSSION

A recent study of rare variants in *CHRNA4* based on a comparison of sequence data for exon 5 for 1209 cases and 1183 controls observed that rare non-synonymous variants were underrepresented in nicotine-dependent subjects, and the authors concluded that such variants might be protective against ND.³³ In the case of R336C, our results clearly show the opposite, namely that the variant encoding this mutation confers risk of ND and its consequences on health, and comparison of the results for the various variants such as P451L and R336C reveals a range of effect sizes (Table 1), suggesting that treating all rare missense variants in *CHRNA4* alike is not a good approach. We note that only one carrier of R336C was observed by Xie *et al.*,³³ a case with ND.

In a follow-up to the aforementioned study,³³ three variants of human *CHRNA4* (encoding for the substitutions R336C, P451L and R487Q) were subjected to detailed functional studies in *Xenopus laevis* oocytes.¹⁸ Electrophysiological studies found changes in the activation by nAChR agonists in channels containing these $\alpha 4$ rare variants, including changes following incubation with nicotine. The authors concluded that sequence variation at *CHRNA4* alters the assembly and expression of human $\alpha 4\beta 2$ nAChRs, resulting in receptors that are influenced to a greater extent by nicotine exposure than channels containing the common $\alpha 4$ variant. Specifically, incubation of oocytes transfected with the $\alpha 4R336C$ variant with nicotine for 24 h had the same effect on nicotine-evoked currents as observed with ACh, revealing a shift in agonist effect that produced a large LS response. Both $\alpha 4R336C$ and $\alpha 4R487Q$ showed a shift from a monophasic, HS relationship to

biphasic activation, with both variants gaining LS activation components. The size of the LS component was particularly large for *a4R336C* (see Figure 4 in McClure-Begley *et al.*¹⁸). Such an effect was not seen for *a4P451L*¹⁸ consistent with the lack of association with FTND for the *P451L* variant (Table 1).

We have presented several lines of evidence for the role of a variant encoding *R336C* in smoking behavior and risk of smoking-related diseases. The variant is rare, but there appears to be a founder effect in Iceland, allowing us to perform association studies with adequate power. First, there is independent and genome-wide significant association between common markers and FTND in the region.⁴ Second, testing a small number of missense variants in *CHRNA4*, we find significant association of the *R336C* variant with FTND and other smoking behavior phenotypes. Third, guided by previous results for *CHRNA5* and smoking-related diseases,¹² we find significant association with four smoking-related diseases, and highly significant association with the early-onset forms. Fourth, functional data show that the *R336C* substitution influences function, shifting the equilibrium following nicotine exposure in favor of a lower-sensitivity form.¹⁸ When taken together, these results allow us to conclude that carriers of this variant have increased risk of nicotine addiction and heavy smoking, which in turn confers considerable risk and lowered age of onset for a number of serious smoking-related diseases through gene-environment correlations³⁴ akin to those previously observed for the *D398N* mutation in *CHRNA5*.^{2,12}

CONFLICT OF INTEREST

Authors whose affiliations are listed as deCODE/Amgen are employees of deCODE/Amgen. The remaining authors declare no conflict of interest.

ACKNOWLEDGMENTS

We thank all the participants in the genetic studies whose contributions made this work possible. This work was supported in part by NIH (NIDA) (R01-DA017932).

AUTHOR CONTRIBUTIONS

The study was designed and results were interpreted by TET, SS, DFG and KS. Phenotype data and Icelandic subject recruitment were coordinated and managed by UT, GB, TET, TR, IJ, AH, SG, SJ, SEM, ThG, TT, TG, HI and HH. LAK, AH, SZ, HJW, WAF, AP and JIM coordinated and managed the lung cancer samples from Holland, Norway, the United States and Spain. IPH, ER, BL, AD, HA and JHP coordinated and managed the COPD samples from the United Kingdom, Sweden and Denmark. GM, UT, PS, DFG, SS, GWR and TET coordinated sequence data analysis, imputation and association analysis. The manuscript was drafted by TET and KS. All authors contributed to the final version of the manuscript.

REFERENCES

- WHO Report on the Global Tobacco Epidemic, 2008: The MPOWER package, WHO, Geneva (2008). Available at: http://www.who.int/tobacco/mpower/mpower_report_full_2008.pdf Accessed February 15, 2016.
- Thorgeirsson TE, Geller F, Sulem P, Rafnar T, Wiste A, Magnusson KP *et al.* A variant associated with nicotine dependence, lung cancer and peripheral arterial disease. *Nature* 2008; **452**: 638–642.
- Thorgeirsson TE, Gudbjartsson DF, Surakka I, Vink JM, Amin N, Geller F *et al.* Sequence variants at *CHRNA3-CHRNA6* and *CYP2A6* affect smoking behavior. *Nat Genet* 2010; **42**: 448–453.
- Hancock DB, Reginsson GW, Gaddis NC, Chen X, Saccone NL, Lutz SM *et al.* Genome-wide meta-analysis reveals common splice site acceptor variant in *CHRNA4* associated with nicotine dependence. *Transl Psychiatry* 2015; **5**: e651.
- Berrettini W, Yuan X, Tozzi F, Song K, Francks C, Chilcoat H *et al.* Alpha-5/alpha-3 nicotinic receptor subunit alleles increase risk for heavy smoking. *Mol Psychiatry* 2008; **13**: 368–373.
- Saccone SF, Hinrichs AL, Saccone NL, Chase GA, Konvicka K, Madden PA *et al.* Cholinergic nicotinic receptor genes implicated in a nicotine dependence association study targeting 348 candidate genes with 3713 SNPs. *Hum Mol Genet* 2007; **16**: 36–49.
- Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T *et al.* Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat Genet* 2008; **40**: 616–622.
- Hung RJ, McKay JD, Gaborieau V, Boffetta P, Hashibe M, Zaridze D *et al.* A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* 2008; **452**: 633–637.
- Timofeeva MN, Hung RJ, Rafnar T, Christiani DC, Field JK, Bickeboller H *et al.* Influence of common genetic variation on lung cancer risk: meta-analysis of 14 900 cases and 29 485 controls. *Hum Mol Genet* 2012; **21**: 4980–4995.
- Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC *et al.* A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet* 2009; **5**: e1000421.
- Lips EH, Gaborieau V, McKay JD, Chabrier A, Hung RJ, Boffetta P *et al.* Association between a 15q25 gene variant, smoking quantity and tobacco-related cancers among 17 000 individuals. *Int J Epidemiol* 2010; **39**: 563–577.
- Ware JJ, van den Bree M, Munafo MR. From men to mice: *CHRNA5/CHRNA3*, smoking behavior and disease. *Nicotine Tob Res* 2012; **14**: 1291–1299.
- Picciozzo MR, Kenny PJ. Molecular mechanisms underlying behaviors related to nicotine addiction. *Cold Spring Harb Perspect Med* 2013; **3**: a012112.
- Rode F, Munro G, Holst D, Nielsen EO, Troelsen KB, Timmermann DB *et al.* Positive allosteric modulation of alpha4beta2 nAChR agonist induced behaviour. *Brain Res* 2012; **1458**: 67–75.
- Henderson BJ, Srinivasan R, Nichols WA, Dilworth CN, Gutierrez DF, Mackey ED *et al.* Nicotine exploits a COPI-mediated process for chaperone-mediated up-regulation of its receptors. *J Gen Physiol* 2014; **143**: 51–66.
- Lester HA, Xiao X, Srinivasan R, Son CD, Miwa J, Pantoja R *et al.* Nicotine is a selective pharmacological chaperone of acetylcholine receptor number and stoichiometry. Implications for drug discovery. *AAPS J* 2009; **11**: 167–177.
- Kuryatov A, Luo J, Cooper J, Lindstrom J. Nicotine acts as a pharmacological chaperone to up-regulate human alpha4beta2 acetylcholine receptors. *Mol Pharmacol* 2005; **68**: 1839–1851.
- McClure-Begley TD, Papke RL, Stone KL, Stokes C, Levy AD, Gelernter J *et al.* Rare human nicotinic acetylcholine receptor alpha4 subunit (*CHRNA4*) variants affect expression and function of high-affinity nicotinic acetylcholine receptors. *J Pharmacol Exp Ther* 2014; **348**: 410–420.
- Cahill K, Stead LF, Lancaster T. Nicotine receptor partial agonists for smoking cessation. *Cochrane Database Syst Rev* 2010; CD006103.
- Heatherton TF, Kozlowski LT, Frecker RC, Fagerstrom KO. The Fagerstrom Test for Nicotine Dependence: a revision of the Fagerstrom Tolerance Questionnaire. *Br J Addict* 1991; **86**: 1119–1127.
- Gudbjartsson DF, Helgason H, Gudjonsson SA, Zink F, Oddson A, Gylfason A *et al.* Large-scale whole-genome sequencing of the Icelandic population. *Nat Genet* 2015; **47**: 435–444.
- Rafnar T, Sulem P, Stacey SN, Geller F, Gudmundsson J, Sigurdsson A *et al.* Sequence variants at the *TERT-CLPTM1L* locus associate with many cancer types. *Nat Genet* 2009; **41**: 221–227.
- Wetzels JF, Kiemeny LA, Swinkels DW, Willems HL, den Heijer M. Age- and gender-specific reference values of estimated GFR in Caucasians: the Nijmegen Biomedical Study. *Kidney Int* 2007; **72**: 632–637.
- Zienoldiny S, Campa D, Lind H, Ryberg D, Skaug V, Stangeland LB *et al.* A comprehensive analysis of phase I and phase II metabolism gene polymorphisms and risk of non-small cell lung cancer in smokers. *Carcinogenesis* 2008; **29**: 1164–1169.
- Stewart CE, Hall IP, Parker SG, Moffat MF, Wardlaw AJ, Connolly MJ *et al.* *PLAUR* polymorphisms and lung function in UK smokers. *BMC Med Genet* 2009; **10**: 112.
- Lundback B, Lindberg A, Lindstrom M, Ronmark E, Jonsson AC, Jonsson E *et al.* Not 15 but 50% of smokers develop COPD?—report from the Obstructive Lung Disease in Northern Sweden Studies. *Respir Med* 2003; **97**: 115–122.
- Ashraf H, Saghir Z, Dirksen A, Pedersen JH, Thomsen LH, Dossing M *et al.* Smoking habits in the randomised Danish Lung Cancer Screening Trial with low-dose CT: final results after a 5-year screening programme. *Thorax* 2014; **69**: 574–579.
- Marchini J, Howie B, Myers S, McVean G, Donnelly P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet* 2007; **39**: 906–913.
- Heatherton TF, Kozlowski LT, Frecker RC, Rickert W, Robinson J. Measuring the heaviness of smoking: using self-reported time to the first cigarette of the day and number of cigarettes smoked per day. *Br J Addict* 1989; **84**: 791–799.
- Chabrol H, Niezborala M, Chastan E, de Leon J. Comparison of the Heavy Smoking Index and of the Fagerstrom Test for Nicotine Dependence in a sample of 749 cigarette smokers. *Addict Behav* 2005; **30**: 1474–1477.
- Thorgeirsson TE, Stefansson K. Genetics of smoking behavior and its consequences: the role of nicotinic acetylcholine receptors. *Biol Psychiatry* 2008; **64**: 919–921.

- 32 Thorgeirsson TE, Stefansson K. Commentary: gene-environment interactions and smoking-related cancers. *Int J Epidemiol* 2010; **39**: 577–579.
- 33 Xie P, Kranzler HR, Krauthammer M, Cosgrove KP, Oslin D, Anton RF *et al*. Rare nonsynonymous variants in alpha-4 nicotinic acetylcholine receptor gene protect against nicotine dependence. *Biol Psychiatry* 2011; **70**: 528–536.
- 34 Plomin R, DeFries JC, Loehlin JC. Genotype-environment interaction and correlation in the analysis of human behavior. *Psychol Bull* 1977; **84**: 309–322.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Supplementary Information accompanies the paper on the Molecular Psychiatry website (<http://www.nature.com/mp>)